(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/31055 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/BE00/00123

(22) International Filing Date: 17 October 2000 (17.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
99870226.0 28 October 1999 (28.10.1999) EP

(71) Applicant *(for all designated States except US)*: **FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX** [BE/BE]; Rue de Bruxelles 61, B-5000 Namur (BE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **REMACLE, José** [BE/BE]; Chemin des Pierres 14, B-5020 Malonne (BE). **ZAMMATTEO, Nathalie** [BE/BE]; Avenue Bourgmestre Jean Materne 202/3, B-5100 Jambes (BE). **DE LONGUEVILLE, Françoise** [BE/BE]; Avenue Jean Materne 110, B-5100 Jambes (BE). **ALEXANDRE, Isabelle** [BE/BE]; Rue du Centre 3, B-5170 Lesve (BE). **HAMELS, Sandrine** [BE/BE]; Allée Saint Hubert 4, B-6280 Loverval (BE).

(74) Agents: **VAN MALDEREN, Joëlle** et al.; Place Reine Fabiola 6/1, B-1083 Brussels (BE).

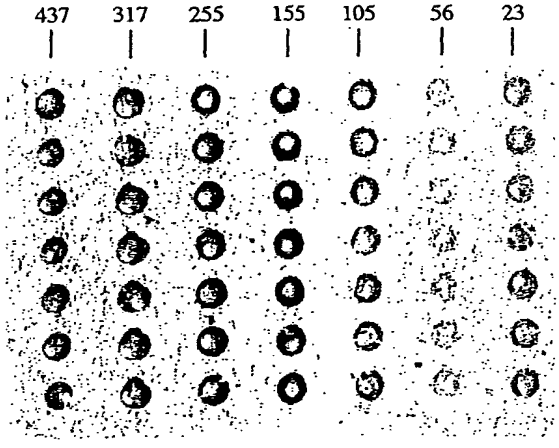
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

[Continued on next page]

(54) Title: METHOD AND KIT FOR THE SCREENING AND/OR THE QUANTIFICATION OF MULTIPLE HOMOLOGOUS NUCLEIC ACID SEQUENCES ON ARRAYS



(57) Abstract: The present invention is related to a new detection and/or quantification method of a nucleotide sequence from other homologous sequences, and being present in a biological sample, comprising the steps of: amplifying or copying at least a part of original nucleotide sequences present in the biological sample into target nucleotide sequences; putting into contact the obtained target nucleotide sequences with corresponding capture nucleotide sequences bounded to an insoluble solid support, characterised in that said capture nucleotide sequences have a (single stranded) length comprised between about 40 and about 400 bases, and in that said capture nucleotide sequences are bounded to the insoluble support according to an array with a density of at least 5 different bound capture nucleotide sequences / cm^2 surface of solid support; and detecting and/or quantifying a signal resulting from the formation of double stranded nucleotide sequences resulting from the hybridisation by complementary base pairing.

WO 01/31055 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND KIT FOR THE SCREENING AND/OR THE QUANTIFICATION OF MULTIPLE HOMOLOGOUS NUCLEIC ACID SEQUENCES ON ARRAYS

Field of the invention

[0001] The present invention is related to a method and kit comprising reagents for simultaneous detection and/or quantification of multiple homologous nucleic acid sequences on arrays.

Background of the invention and state of the art

[0002] The development of miniaturisation, especially the array technology, allows the simultaneous detection of a great number of nucleotide sequences present in biological samples. The array contains on its surface, series of discrete regions bearing capture nucleotide nucleotide sequences able to bind by hybridisation corresponding target nucleotide sequences. If the last ones are labelled, a signal can be detected and measured at the binding location and its intensity gives an estimation of the amount of target sequences present in the sample.

[0003] It is known that the presence of a particular oligonucleotide sequence at a specific location on a support can be obtained by using masks at each step comprising the addition of a new nucleotide on a growing oligonucleotide in order to obtain the desired sequence. This mask technique is derived from the photolithographic technology and is coupled with the use of photoprotective groups, which are released before a new nucleotide is

added. The obtained array can contain several thousands of discrete domains, each one bearing a specific oligonucleotide as capture nucleotide sequence for the possible detection of thousands of target sequences. Only small oligonucleotides are present on the surface, and are used for sequencing or identifying sequences in a pattern of positive spots corresponding to specific oligonucleotides present on the array. The identification of the target sequence is made by comparison of said pattern with a reference. The documents WO97/29212 and WO98/28444 describe a similar technique with the use of capture nucleotide sequences comprising less than 30 nucleotides allowing an analysis of two different sequences that may differ by a single nucleotide. Small capture nucleotide sequences (between 10 and 20 nucleotides) are preferred since the discrimination between two oligonucleotides differing in one base is higher, when their length is smaller.

[0004] However, for longer target fragments (like the cDNA copy of mRNA), several fragments can recognise the same nucleotide sequence and the rate of hybridisation is lower. Therefore, the fragments are cut into smaller species and the method should require the use of numerous capture nucleotide sequences in order to obtain a pattern of signals which attest the presence of a given gene (WO97/10364 and WO97/27317). In this case, short capture sequences are used, preferably about 45 nucleotides, and in practice, between 10 and 25 nucleotides.

[0005] The document US-5,605,662 describes the binding of full nucleotide sequences on plots of electronic based chips, each of the plots bearing a specific oligonucleotide sequence. A target sequence is spotted on the chips and specific small labelled nucleotides are then incubated in order to determine their binding to the

target. A change in the plot charge allows a discrimination of the different nucleotide differing in one base, which allows a specific SNP analysis (Gilles et al. 1999, Nature Biotechnology 17, p. 365 (1999)).

[0006] The document WO94/22889 describes electronic chips made of preconstructed oligonucleotides bearing a pyrol group used in an electropolymerisation step and allows the formation of a conducting polymer. This method is particularly suitable for the preparation of electronic chips bearing small capture nucleotide sequences.

[0007] The approaches using either small or long capture nucleotide sequences fixed on array, are not optimal conditions when applied for diagnostic purpose, especially, when different sequences having a high homology between them should be detected and possibly quantified at the same time upon the same microarray.

Aims of the invention

[0008] The present invention aims to provide a new method to improve the detection and possibly the quantification of nucleotide sequences, preferably multiple homologous sequences, coming from different or the same organism, by hybridisation on single stranded capture nucleotide sequences, and which does not present the drawbacks of the state of the art.

Summary of the invention

[0009] The inventors have discovered that it is possible to provide a sensitive detection method (possibly combined with a quantification) of multiple nucleotide sequences upon an array, even if said multiple sequences are homologous and are present simultaneously in the sample submitted to the analysis. Until now, either the sensitivity was low or null or there was cross-reactivity

between the different (but homologous) target sequences on the same capture nucleotide sequences.

[0010] One unexpected observation was the result obtained when a given target sequence was incubated with an array bearing corresponding capture nucleotide sequences of various length (example 2 and figure 1). With a target sequence made of a double stranded amplicon (just denatured before the experiment), either the use of too long or too small capture nucleotide sequences gave a lower signal (and sometimes a much lower signal) compared to the one obtained when capture nucleotide sequences have the same length as the target sequence. The highest sensitivity was obtained with capture nucleotide sequences with moderate length and a size similar to the target amplicon. For very long target amplicons, the situation is much more complicated (reassociation of the amplicons, secondary structures, ...) and depends on the nature of the sequence. In sandwich hybridisation of long target amplicons upon microplates, optimal capture nucleotide sequences have been found to be between 50 and 500 bases. (WO98/11253).

[0011] The present invention is related to an (improved in sensitivity and specificity compared to known techniques) detection and/or quantification method of a nucleotide sequence from at least 4 others homologous sequences, possibly present in a biological sample, by the identification of a portion of its sequence, said method comprising the steps of:

- possibly extracting original nucleotide sequences from the biological sample,
- amplifying or copying at least a part (or portion) of the original nucleotide sequences into target nucleotide sequences, each possible target nucleotide sequence being located at a different location of the sequence.

- possibly denaturing the double stranded target nucleotide sequences into single target nucleotide sequences,
- possibly cleaving said sequences into smaller target nucleotide sequences having a length higher than about 40 and lower than about 400 base pairs (or at a preferred size described hereafter) by various methods (restriction enzymes, addition of NaOH, etc.),
- possibly labelling said target nucleotide sequences,
- putting into contact the obtained target nucleotide sequences with corresponding capture nucleotide sequences bounded (linked or fixed) to an insoluble solid support, said capture nucleotide sequences having a (single stranded) length comprised between about 40 and about 400 bases, preferably between about 50 and about 350 bases, more preferably between about 100 and about 300 bases, even more preferably between about 120 and about 200 bases, said capture nucleotide sequences being bound to the insoluble solid support according to an array with a density of at least 5 different bound capture nucleotide sequences/cm^2 surface of solid support.

[0012] The method comprises also the step of detecting and/or possibly quantifying a signal (preferably from the label of the target nucleotide sequences) resulting from the formation of double stranded nucleotide sequences resulting from the hybridisation (of the capture and target nucleotide sequences) by base pairing.

[0013] The labelling (with known techniques from the person skilled in the art) is preferably also obtained upon the amplified sequence previously to the denaturation (if the method comprises an amplification step).

[0014] The method according to the invention may also further comprise the step of correlating the signal of detection to the identification (location) of the specific target nucleotide sequence (one spot = 1 specific sequence) without comparing the hybridisation pattern obtained with a standard pattern as described in the state of the art (WO97/02102), to the presence of specific microorganisms, or genetic characteristics (polymorphism, genetic diseases, diagnostic and monitoring of cancer, etc.) for a patient (from which the biological sample has been obtained). The biological sample can be any culture medium wherein microorganisms or pollutants are present, or an extract obtained from a plant or an animal (including a human) organ, tissue, cell or biological fluid.

[0015] The various steps of the method according to the invention can be done with various means well known by the person skilled in the art and described in the literature.

[0016] The method according to the invention can be performed by using a specific diagnostic and/or quantification kit comprising at least an insoluble solid support upon which capture nucleotide sequences are disposed (preferably bounded to the solid support by a covalent link) according to an array with a density of at least 5 different bound capture nucleotide sequences / cm^2 surface of the insoluble solid support, said capture nucleotide sequences having a length comprised between about 40 and about 400 bases or preferably a length as above-described.

[0017] In the method and kit according to the invention, the density of the capture nucleotide sequences upon the array of the solid support can be increased, for instance by having more than 10, 20, 50, 100, 1000, 10000

or more capture nucleotide sequences/cm^2 surface of solid support. In the preferred embodiment, each capture nucleotide sequence is located in a specific area of the support. However, several capture nucleotide sequences can also be present in the same area to obtain specific information.

[0018] The kit according to the invention may also incorporate various media or devices for performing (preferably automatically) the method according to the invention. Said kit can be included in an automatic device, such as a high throughput screening apparatus for the detection and/or the quantification of multiple nucleotide sequences present in a biological sample to be analysed. Said kit or apparatus can be adapted for performing all the steps or only several specific steps of the method according to the invention.

[0019] In the method, kit or apparatus according to the invention, the length of the single stranded capture nucleotide sequences is preferably identical to the length of the target nucleotide sequences to be detected and/or quantified or may differ, preferably by less than 50% in total length, more preferably less than 30% in total length, even more preferably less than 10% in total length.

[0020] The method, kit or apparatus according to the invention are suitable for the detection and/or the quantification of target nucleotide sequences which are made of DNA or RNA, including sequences which are partially or totally homologous upon their total length.

[0021] The method according to the invention can be performed even when the different original nucleotide sequences present between them an homology greater than 30%, greater than 60% or even greater than 80% or differ by few bases.

[0022] In the method, kit or apparatus according to the invention, the capture nucleotide sequences are advantageously covalently bounded (or fixed) upon the insoluble solid support, preferably by one of their extremities as described hereafter.

[0023] With the method according to the invention, the yield of hybridisation is advantageously greater than 10%, 50%, 70%, 80% or 90% or can achieve almost 100%.

[0024] The method, kit and apparatus according to the invention may comprise the use of other bounded capture nucleotide sequences (i.e. by allowing an hybridisation with a standard nucleotide sequence used for the quantification, with a consensus sequence of different micro-organisms strains or with a sequence allowing a detection of a possible an antibiotic resistance by micro-organisms) or other non-homologous sequences, said other capture nucleotide sequences having possibly a length higher than 400 bases and being also bounded upon the insoluble solid support (biochip), preferably in the array made with the other bounded capture nucleotide sequences.

[0025] The solid support according to the invention can be or can be made with materials selected from the group consisting of glasses, electronic devices, silicium or plastic support, compact discs, filters, metallic supports or a mixture thereof. Advantageously, said solid support is a single glass plate which may comprise additional means (barcodes, markers, etc.) or media (coating, etc.) for improving the method according to the invention.

[0026] The amplification step(s) used in the method according to the invention is advantageously obtained by well known amplification protocols, preferably selected from the group consisting of PCR, LCR, CPR, NASBA, ICR or Avalanche DNA techniques.

[0027] Advantageously, the length of the target nucleotide sequences to be detected is determined by the conditions of the above-identified amplification protocols determined by the use of specific primers (and possibly blocking nucleotide sequences) for a retro-transcription of the 3' or 5' end of the original biological nucleotide sequences to be detected and/or quantified, especially if it is a RNA sequence. Advantageously, said RNA sequences are 16S and 23S rRNA sequences or 18S and 28S rRNA sequences.

[0028] In a preferred embodiment of the present invention, the target nucleotide sequences to be copied or amplified are obtained from different parts or portions of the corresponding (homologous) DNA or RNA original biological nucleotide sequence.

[0029] In a preferred embodiment of the invention, the primers used for the amplification preferably bear at their 3' end a base specific of one of the homologous sequences.

[0030] According to a further aspect of the present invention, the method, kit or apparatus according to the invention is advantageously used for the detection and/or the quantification of different Staphylococci species or variants, preferably the Staphylococcus aureus, the Staphylococcus epidermidis, the Staphylococcus saprophyticus, the Staphylococcus hominis or the Staphylococcus haemolyticus present together or separately in a biological sample, said detection being obtained by detecting the genetic variants of the FemA gene in said different species, preferably by using specific locations in the FemA genetic sequence, as described in the document WO99/16780 incorporated herein by reference.

[0031] Preferably, the primers and the specific portions of said FemA sequence used for obtaining amplified products are the ones described hereafter in the examples.

[0032] The method according to the invention may also comprise means for obtaining a quantification of target nucleotide sequences by using a standard nucleotide sequence (external or internal standard) which can be brought into contact with the capture nucleotide sequences bounded upon the array of the solid support in the same conditions as said target nucleotide sequences (possibly after amplification or copying). Said method comprises a step of quantification of a signal resulting from the formation of a double stranded nucleotide sequence formed by complementary base pairing between the capture nucleotide sequences and the standard nucleotide sequences and a step of a correlation analysis between the signal resulting from the formation of said double stranded nucleotide sequence and the signal resulting from the double stranded nucleotide sequence formed by complementary base pairing between capture nucleotide sequences and target nucleotide sequences in order to quantify the presence of the original nucleotide sequence to be detected and/or quantified in the biological sample (see also the document WO98/11253 incorporated herein by reference). Said standard nucleotide sequence (external and/or internal standard) is advantageously included in the kit or apparatus according to the invention, possibly with the media and device(s) necessary for performing the different steps according to the invention, such as the hybridisation and culture media, polymerases, enzymes, standard sequences and labelling molecules.

[0033] The present invention will be described in details in the following non-limiting examples in reference to the enclosed figures.

Brief description of the drawings

[0034] Fig. 1 shows the influence of the capture nucleotide sequence length on the yield of hybridisation of target amplicons. The target sequences were 155 bases long and 100 fmoles were incubated on a chips containing single stranded capture nucleotide sequences going from 23 to 437 bases long. The length of the capture nucleotide sequence is shown on the figure.

[0035] Fig. 2 schematically represents the FemA detection of 5 different species of Staphylococci. The locations of the 5 pairs of primers used for specific amplification of one of the 5 different sequences belonging to the 5 Staphylococci species are shown on the FemA sequence. A consensus sequence is also amplified by using 2 primers common to all Staphylococci species. The labelled amplified sequences are then hybridised on the chips.

[0036] Fig. 3 schematically represents the detection of rRNA by the copy of small portions of the sequence either at its 5' end using one starting nucleotide sequence or along its sequence using both a starting nucleotide sequence and a blocking nucleotide sequence. The small copied sequences are then hybridised on the array of the chips.

Definitions

[0037] The term "nucleoside triphosphate" refers to nucleosides present in either DNA or RNA and thus includes nucleosides which incorporate adenine, cytosine, guanine, thymine and uracil or other modified bases (i.e. 8-azaguanine and hypoxanthine) as bases, the sugar moieties being deoxyribose or ribose.

[0038] The term "nucleotide" as used herein refers to nucleosides present in nucleic acids (either DNA or RNA) compared with the bases of said nucleic acid, and includes

nucleotides comprising usual or modified bases as above described. References to nucleotide(s), oligonucleotide(s) and the like include analogous species wherein the sugar-phosphate backbone is modified and/or replaced, provided that its hybridisation properties are not destroyed (the backbone may be replaced by an equivalent synthetic peptide, Peptide Nucleic Acid (PNA)).

[0039] The primer sequence need not reflect the exact sequence of the template to be amplified or copied provided that under hybridising conditions the primers can be used in a genetic amplification. Mismatched bases can be introduced into the primer sequence to provide altered hybridisation introduced into the primer sequence to provide altered hybridisation stringency (amplification or copying of the same regions of homologous sequences by the same primers).

[0040] The terms "capture or target nucleotide sequences, target nucleic acid, hybridising specifically to, background, quantifying" are defined as in WO97/10365, incorporated herein by reference.

[0041] "Homologous sequences" mean nucleotide sequences having the same nucleotide at corresponding positions. They are generally defined as a minimum of homology (or sequence identity) between sequences wherein the percentage of identical nucleotides after the sequences has been optimally aligned taking into account additions or deletions like gaps (for sequences of a given gene, present in genetically different sources like different organisms or for proteins or enzymes of the same family). The degree of homology (or sequence identity) can vary a lot as homologous sequences may be only in one or several portions or all along the complete sequences. The parts or portions of the sequences that are identical in both sequences are

said conserved. The sequences showing a high degree of invariance or homology are said to be highly conserved.

[0042] Methods of alignment of sequences are based on local homology algorithms which have been computerised and are available as for example (but not limited to) Clustal®, (Intelligenetics, Mountain Views, California), or GAP®, BESTFIT®, FASTA® and TFASTA® (Wisconsin Genetics Software Package, Genetics Computer Group Madison, Wisconsin, USA) or Boxshade®.

Detailed description of the invention

[0043] The amplification of target sequences of 100 to 200 bases and hybridisation on capture nucleotide sequences comprised between 100 and 200 bases allows to solve many problems posed by the differential diagnostic of homologous sequences. However smaller lengths (especially for capture nucleotide sequences) going down to 40 bases and longer length going up to 400 bases can also be used (mostly limited by the amplification step which needs the use of two primers and some sequences to be copied). The reduction of the capture nucleotide sequence for a given target sequence leads to a decrease in sensitivity, but an increase in specificity. Longer target sequences limit the possibility to design the various targets amplicons on the same sequence as shown in Fig. 2. It also reduces specificity and can lead to secondary structures decreasing the yield of hybridisation.

[0044] Target DNA sequences are amplified by classical methods like PCR using primers so that small fragments are produced which are then used for hybridisation on an array bearing the corresponding capture nucleotide sequences. RNA can be copied and retro-transcripted in cDNA and amplified the same way if necessary.

[0045] When homology between the sequences to be detected in the sample is not too high (typically between 30 and 60%), the homologous sequences are amplified or copied (for instance by using the same primers) and the discrimination is obtained on the array using moderate length capture nucleotide sequences as proposed in the present invention. The use of the same primers is possible given that some parts of the sequences are similar or even identical on a certain distance and it facilitates the amplification step. However since the sequences are also rather different makes possible to capture them specifically on moderate length capture nucleotide sequence. The fact that both target and capture nucleotide sequences are of similar length makes the detection sensitive.

[0046] However when several sequences with high homology (for instance higher than 60%) have to be analysed the discrimination by hybridisation on capture nucleotide sequences only is not always sufficient (cross-reactions between two homologous sequences on the same capture nucleotide sequence). Therefore, different small fragments of the target sequences are either amplified or copied. These fragments are located in different part of the sequence for the different targets. Some parts of these locations can be overlapping so that the locations are not totally but only partially distinct. This situation is compatible with the invention as long as there are no cross reactions on the capture nucleotide sequences by the different target nucleotide sequences. These small fragments are then detected on their specific capture nucleotide sequences of moderate length. High sensitivity is obtained throughout the use of capture nucleotide sequences of moderate length similar to the sequences to detect. The specificity is obtained first during the

amplification using specific primers for each sequence to be amplified or copied and secondly on the microarray. For a sample analysis, the different primers are used together in a multiplex PCR amplification before analysis on the biochips. Preferably, one may use primers specific for each target having also at their 3' end a base specific of the sequence to detect so that (in appropriate conditions) only the target sequence bearing the corresponding base will be copied and amplified by the polymerase.

[0047] In an alternative, a semi-consensus multiplex PCR was tested by using a primer common for all sequences and a second primer specific of each homologous sequence. Specific capture nucleotide sequences of moderated length were selected for each sequence in a different part of the overall target sequence.

[0048] In another alternative, which can be combined with the previous mentioned ones, multiple capture nucleotide sequences which recognise different parts of a same target are fixed to the array, which contain two or three capture nucleotide sequences binding to different parts of a given amplified sequence. The resulting signals for each of these spots are in a given ratio for a given target sequence, while it is different for another sequence which partly cross-reacts as the homology between two sequences is not homogeneous (with some parts being very different while others are more conserved).

[0049] The assay for RNA present in multiple copies in cells and in large amount like the 16s or 23s rRNA in procaryote cells or the 18s or 28s rRNA in eucaryote cells does not require necessarily amplification. The detection based on their hybridisation on microarray is particularly well suited since they show a high degree of homology between different species. It is possible to copy one specific part of the RNA by using a primer to start the

copy and a blocking nucleotide sequence which binds to the RNA and stops the copying by the reverse transcriptase (example 5). This blocking oligonucleotide has its 3' end blocked so that it can not be used by the transcriptase to start a copy. The most simple 3' block being a deoxy 3' carbon, but others like the presence of a pyrophosphate at the 3' carbon or a 2',3'-dideoxycarbon are also working. The 5' end of this blocking oligonucleotides is also modified like for example by a NH_2 group in order to block transcriptase. The primer can also bear at its 3' end a base specific of a given sequence, in order to make a more specific copy. Therefore, different specific parts of the rRNA of different organisms can be copied making their detection on the array unequivocal since not only their sequence, but also their location will be specific.

[0050] The same copy of the RNA can be obtained with messenger RNA (mRNA) which can be transcripted in cDNA and amplified by PCR if necessary.

[0051] The development of chemistry allows the covalent fixation to glass of nucleotides bearing a specific functional group (Lamture et al., Nucleic Acid Res. 22, pp. 2121-2125 (1994)). By using aminoterminal groups present on the oligonucleotides, it is possible to obtain a covalent binding of nucleotides upon aldehyde groups present on the array. The capture nucleotide sequences are either chemically synthesised or produced by PCR (amplicons) and bounded to a functionalised glass by a robot and thereafter rendered single stranded. The capture nucleotide sequences have complementary sequences related to the target DNA to detect and have a similar, if not identical length. Difference of 50 % in the length still gives a high sensitivity binding. The array is constructed with an appropriated automate which deposits the capture

nucleotide sequences at a given location which delimitates a spot. Triplicate of each capture nucleotide sequence plus the negative and positive controls (and possible standard sequences) are used. The array contains typically between 20 and 100 spots, but arrays going to 400 spots, 1000 or 10000 spots or more are possible. An array with 400 spots per cm^2 is obtained with pins of 0.2 mm at low cost. As the capture nucleotide sequences are present in a sufficient number, it allows the lecture of these spots with a great resolution with known detection apparatus.

[0052] Other supports like filters either of nitrocellulose or nylon are also proposed for arrays. The binding sites of the DNA on these filters is however at random along the sequence and it is impossible to predict the size of the available single stranded sequence which is then available for hybridisation.

[0053] Binding of oligonucleotides sequences can be obtained through covalent or non-covalent binding on these supports in a direct or indirect reaction. One common method is to spot the capture nucleotide sequences on a surface where polylysine has been attached on glass or plastic through the binding of biotinylated oligonucleotides on streptavidin coated surfaces or through the use of proteins with binding affinity for oligonucleotides (EP-A-0491059).

[0054] Gel layers (US-A-5,552,270 and EP-0535242), copolymerisation of acrylamide with vinyl bearing oligonucleotides (US-A-5,736,257) or other supports, which can be activated for covalent fixation of oligonucleotides, can be used for arrays. Plastic like polycarbonate as present on CD was activated in order to fix capture nucleotide sequences and be used as a bio-CD array (WO99/35499).

[0055] Hybridisation between two DNA or RNA chains is a complex process which is initiated by the binding of a few (4-5) nucleotides which recognised themselves in a specific way and once they are bound a very fast process thermodynamically favourable elongation of the binding occurs along the sequence. The binding is thus dependent both of kinetic and thermodynamic parameters and experimental conditions can be adapted in order to modify both of them. Temperature accelerates the kinetic process but there is an optimum for the temperature used to obtain a maximum binding of the target. The salt concentration modifies the stringency conditions: more salts present in the solution, more easy will be the binding of the two chains. Preferred conditions are obtained with targets and capture nucleotide sequences of the same size and of moderate length.

[0056] A very good yield of capture of the target on the capture nucleotide sequences can be obtained even if the targets are double stranded like after PCR amplification. The hybridisation of the target DNA is a competitive reaction: the target strand can hybridise on the fixed capture nucleotide sequence but can reassociate in solution with its complementary strand. The reaction in solution is always kinetically favourable due to the free movement of molecules in solution. The rate of reaction is proportional to the square root of the length of the shorter strand. When capture nucleotide sequences have almost similar length as the target, the two reactions are independent of the length, since they are the same or similar. An optimum number of the same capture nucleotide sequences has been fixed per spot so that the high number of capture nucleotide sequences compensates for the loss obtained by the diffusion constrains. Moderated length

capture nucleotide sequences disposed at a certain distance from the surface lower the diffusion effect.

[0057] Other components of reactive solutions have also to be incorporated like buffer(s), detergent(s), DMSO or the addition of non specific DNA like salmon DNA.

[0058] After hybridisation on the array, the target sequences can be detected by current techniques. Without labelling, preferred methods are the identification of the target by mass spectrometry now adapted to the arrays (US-A-5,821,060) or by intercalating agents followed by fluorescent detection(WO97/27329 or Fodor et al., Nature 364, p. 555 (1993)).

[0059] The labelled associated detections are numerous (see review in W0 97/27317). They are obtained using either already labelled primer or by incorporation of labelled nucleotides during the copying or amplification step. A labelling can also be obtained by ligating a detectable moiety onto the RNA or DNA to be tested (a labelled oligonucleotide ligated at the end of the sequence by a ligase). Fragment of RNA or DNA can also be incorporate labelled nucleotides at their 5'OH or 3'OH ends using a kinase, a transferase or a similar enzyme.

[0060] Labels like fluorescent nucleotide sequences like Cy3, Cy5 and Cy7 are suitable for analysing an array by using commercially available array scanners (General Scanning of Genetic Microsystem). Radioactive labelling, cold labelling or labelling with small molecules recognised thereafter by specific ligands (streptavidin or antibodies) are common methods. The resulting signal of target fixation on the array is either fluorescent, colorimetric, diffusive, electroluminescent, bio- or chemiluminescent, magnetic, electric like impedometric or voltametric (US-A-5,312,527). The two preferred embodiments of the invention are the fluorescent detection or the gold

labelling of the bound target in order to obtain a precipitate or silver staining which is then easily detected and quantified by a scanner (EP-99870106.4).

[0061] The signal obtained for each spot is recorded and the mean of the signal is calculated for identical capture nucleotide sequences. In practice at least two and preferably three to five identical spots are present on each array in order to correct for variation which can occur at any step of the process. The background value is identified either in the part of the array which has no capture nucleotide sequence or on spots bearing non specific capture nucleotide sequence (negative control). A positive control is preferably added (a DNA sequence which is added to the hybridisation solution and in which capture nucleotide sequence is present at least on one spot of the array), to test for hybridisation step, solutions and conditions used and detection. Different positive nucleotide sequences present at various concentrations can also be added to the sample in order to obtain a reference curve for the signal. The various signals of the spots can then be compared to this reference curve.

[0062] Quantification takes into account the hybridisation yield and detection scale on the array (which is identical for target and reference sequences) and the extraction, the amplification (or copying) and the labelling steps. Internal standard are used in quantification by the measurement of the target sequence compared to a given sequence (reference) and to which the other values will be compared.

[0063] External standard can also be added to the sample for the quantification. If PCR is used, an internal standard contains at its extremities the same sequences as the target in order to be amplified by the same primers. It can also be of the same length, has the same GC content or

even have a large part of its sequence identical to the target in order to be really competitive during the amplification step.

Examples

Example 1: Detection of target nucleotide sequences on an array

[0064] The protocol described by Schena et al (Proc. Natl Acad. Sci. USA 93, 10614 (1996)) was followed for the grafting of aminated DNA to aldehyde derivatised. The long aminated capture nucleotide sequences (100-400 bases) were spotted at a concentration of 150 nM while the small oligonucleotides were at 450 nM. The capture nucleotide sequences were printed onto the silylated microscopic slides with a home made robotic device, using 250 μm pins from Genetix (UK) and silylated (aldehyde) microscope slides from Cell associates (Houston, USA). The spots have 400 μm in diameter and the volume dispensed is about 1 nl. Slides were dried at room temperature and stored at 4 °C until used.

[0065] 5 μl of hybridisation solution were loaded on glass slides bearing the capture nucleotide sequences. This mixture contained : SSC2X, SDS 4%, salmon sperm DNA 100 μg/ml, 2 nM biotinylated CMV amplicons of 437 bp and 10 nM of biotinylated target amplicons. Microarrays were covered with coverslips prewashed with ethanol 100%. Slides were denatured at 95 °C for 5 min. The hybridisation was carried out at 65° for 2 h. Samples were washed 4 times with Maleic buffer 10 mM pH 7.5, NaCl 15 mM, Tween 0.1%.

[0066] The glass samples were incubated 45 min at room temperature with 800 μl of streptavidin labeled with colloidal gold. After washing, the presence of gold served for catalysis of silver reduction using the staining

revelation solution (Sigma St Louis, Mi). The slides were dried before being store at room temperature and analysed using a micro array reader.

Example 2: Comparison of the sensitivity obtained for hybridisation of a target sequence of medium size (155 bp) on capture nucleotide sequences of various length

[0067] The protocols for capture nucleotide sequences immobilisation and silver staining detection are described in example 1. The capture nucleotide sequences and target DNA were obtained by amplification of CMV sequence by PCR using the following primers:

Primer	Sequence (5' -> 3')	Length (bp)	Location of Primer (d)
Target[a]			
MIE-4	CCAAGCGGCCTCTGATAACCAAG		2223-2245
MIE-5c	CAGCACCATCCTCCTCTTCCTCTGG	437	2657-2633
MIE-6c	GGCGATGGCCCGTAGGTCATCCA	155	2374-2352

Capture nucleotide sequence[b]			
MIE-4c	CCAAGCGGCCTCTGATAACCAAG		2223-2245
MIE-6a	GCGGCGCTTCATTACACTGATAAC	56	2275-2252
MIE-6b	CGGCCCCCAGAATGTACTGGGCA	105	2324-2302
MIE-6c	GGCGATGGCCCGTAGGTCATCCA	155	2374-2352
MIE-6d	GTACAGGGGACTCTGGGGGTGAC	255	2474-2452
MIE-6[e]	CTGCTCACTTTCTTCCTGATCACTG	317	2536-2560
MIE-5c	CAGCACCATCCTCCTCTTCCTCTGG	437	2657-2633

(a) The use of primer MIE-4 with respectively primer MIE-5c and MIE-6c resulted in target DNA of 437 and 155 bp

(b) The use of primer MIE-4 with respectively primer MIE-6a, MIE-6b, MIE-6c, MIE-6d, MIE-6e and MIE-5c resulted in capture nucleotide sequences of 56, 105, 155, 255,

317 and 437 bp.

(c) MIE-4 was aminated at 5' end.

(d) Expressed as nucleotide number counted sequentially within the MIE gene.

[0068] Plasmid pAT153-E (containing the exon 4 of the MIE gene of HCMV DNA AD169 strain) was amplified by PCR, in a 100 µl volume containing 1.5 mM $MgCl_2$, 10 mM Tris pH 8.4, 50 mM KCl, 1 µM of each primer, 100 µM of each dNTP, 2.5 U of Taq DNA polymerase Gold and 10 ng of plasmid pAT153-E. Samples were first denatured at 94 °C for 10 min to activate the polymerase. Then 40 cycles of amplification were performed consisting of 30 sec at 94 °C, 30 sec at 65 °C and 30 sec at 72 °C and a final extension step of 10 min at 72 °C. Water controls or 100 copies of plasmid DNA were used respectively as negative or positive controls of the amplification. PCR of target DNA also includes 100 µM of biotin-16-dUTP.

[0069] The hybridisation step was carried out at 65 °C for 2 h in the presence of 100 fmoles of biotinylated target DNA.

Example 3: Influence of the capture nucleotide sequence length on the yield of fixation of long target amplicons (437 bp)

[0070] The experiment was conducted as described in example 2, with a target of 437 bp, obtained by amplification of the Plasmid pAT153-E by using the primers MIE-4 and MIE-5c. Hybridisation was conducted on the biochips of example 2.

Example 4: Detection of FemA sequences from different bacterial species on the same array

[0071] The FemA genes corresponding to the different Staphylococcus species were amplified separately by multiplex PCR using the following primers:

S. aureus 1 :	5' CTTTTGCTGATCGTGATGACAAA 3'
S. aureus 2 :	5' TTTATTTAAAATATCACGCTCTTCG 3'
S. epidermidis 1 :	5' TCGCGGTCCAGTAATAGATTATA 3'
S. epidermidis 2 :	5' TGCATTTCCAGTTATTTCTCCC 3'
S. haemolyticus 1 :	5' CTATGGTATTAGCGGTAATTTTAG 3'
S. haemolyticus 2 :	5' TTTAATCTTTTTGAGTGTCTTATAC 3'
S. saprophiticus 1 :	5' TAAAATGAAACAACTCGGTTATAAG 3'
S. saprophiticus 2 :	5' AAACTATCCATACCATTAAGTACG 3'
S. hominis 1 :	5' CGACCAGATAACAAAAAAGCACAA 3'
S. hominis 2 :	5' GTAATTCGTTACCATGTTCTAA 3'

[0072] The location of these primers and their specificity for the different Staphylococcus are presented in figure 2.

[0073] The multiplex PCR was performed in a final volume of 50 µl containing: 1.5 mM $MgCl_2$, 10 mM Tris pH 8.4, 50 mM KCl, 0.8 µM of each primer, 50 µM of each dNTP, 50 µM of biotin-16-dUTP), 1.5 U of Taq DNA polymerase Biotools, 7.5% DMSO, 5ng of plasmid containing FemA gene. Samples were first denatured at 94 °C for 3 min. Then 40 cycles of amplification were performed consisting of 30 sec at 94 °C, 30 sec at 60 °C and 30 sec at 72 °C and a final extension step of 10 min at 72 °C. Water controls were used as negative controls of the amplification. The sizes of the amplicons obtained using these primers were 116 bp for *S. saprophiticus*, 128 bp for *S. aureus*, 118 bp for *S. hominis*, 162 pb for *S. epidermidis* and 160 bp for *S. haemolyticus*.

[0074] The sequences of the capture nucleotide sequences were the same as the corresponding amplicons and single strands.

[0075] Protocols for capture nucleotide sequences immobilisation, hybridisation and silver staining detection are the one described in example 2.

Example 5: Detection of 16S rRNA from different bacteria by copying of a small portion of the sequence

[0076] A specific sequence of the 16S rRNA sequence of three different bacteria (*E.coli*, *Bacteroides distasonis* and *Bifidobacterium longum*) was copied using both a starting nucleotide sequence and a blocking nucleotide sequence which hybridises on the 16S rRNA and stops the reverse transcription. The copy of a specific sequence was done on 2µg of total RNA extracted from different bacteria. The following sequences were used as starting and blocking nucleotide sequences:

- for *E.coli*

  Starting nucleotide sequence

  5' CTCTGAAAACTTCCGTGGATG 3'

  Blocking nucleotide sequence

  5' GTATCTAATCCTGTTTGCTCCCCACGCT 3'

- for *B.distasonis*

  Starting nucleotide sequence

  5' TACGATCCATAGAACCTTCATCCC3'

  Blocking nucleotide sequence

  5' CCTGCTTCATGCGGTATTAGTCCGAC3'

- for *B.longum*

  Starting nucleotide sequence

  5' CCACCGTTACACCGGGAA3'

  Blocking nucleotide sequence

  5' CTCTCGCTTGCTCCCCGATA3'

[0077] The starting nucleotide sequence for *E.coli*, *Bacteroides distasonis* and *Bifidobacterium longum* are located respectively at position 995-1016, 420-443,658-675 on the 16S rRNA.

[0078] The blocking nucleotide sequence for *E.coli*, *Bacteroides distasonis* and *Bifidobacterium longum* are located respectively at position 778-797, 172-198, 441-460 on the 16S rRNA. The length of the synthetized cDNA are respectively of 240 bases, 273 bases and 236 bases.

[0079] In a nuclease free microtube, 0.5 µg of the starting nucleotide sequence and 2 µg of the blocking nucleotide sequence were added to 2 µg total RNA extracted from bacteria. Nuclease free water was added to a final volume of 15 µl. The reverse transcription was conducted by adding the following components to the annealed nucleotide sequence /template : 5 µl of 5X AMV RT Buffer (250 mM Tris-HCl pH 8.3, 250 mM KCl, 50 mM MgCl2 50 mM DTT and 2.5 mM Spermidine), 40 units of Rnasin ribonuclease inhibitor(Promega, Madison, US), 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 0.65 mM dTTP, 0.35 mM biotin dUTP and 30 units of AMV RT(Promega, Madison, US). The final volume was adjusted to 25 µl with nuclease free water. The reaction mixture was mixed gently and incubated for 60 min at 60 °C.

[0080] The single strands cDNA obtained from the different bacteria were hybridised and detected on the array as in example 21.

[0081] The results showed a specific hybridization of the cDNA for each of three bacteria on their respective capture nucleotide sequence. There was no cross-reactions since their values were at the level of the backgroung.

CLAIMS

1. Detection and/or quantification method of a nucleotide sequence from at last 4 other homologous sequences, and being possibly present in a biological sample, comprising the steps of:

- amplifying or copying at least a part of an original nucleotide sequence present in the biological sample into target nucleotide sequences, each target nucleotide sequence being located at a different location of the original nucleotide sequence;
- putting into contact the obtained target nucleotide sequences with corresponding capture nucleotide sequences bounded to an insoluble solid support,
- characterised in that said capture nucleotide sequences have a (single stranded) length comprised between about 40 and about 400 bases, and in that said capture nucleotide sequences are bound to the insoluble support according to an array with a density of at least 5 different capture nucleotide sequences / cm2 surface solid support, and
- detecting and possibly quantifying a signal resulting from the formation of double stranded nucleotide sequences resulting from their hybridisation by complementary base pairing.

2. The method of claim 1, wherein the length of the capture and target nucleotide sequences is comprised between about 50 and about 300 bases, preferably between about 100 and about 200 bases.

3. The method of claim 1 or 2, wherein the length of the capture nucleotide sequences differs less than 50% to the length of the target nucleotide sequences.

4. The method of any of the preceding claims 1 to 3, wherein the different target nucleotide sequences

to be detected present an homology between each other higher than 30%, preferably higher than 60%, more preferably higher than 80%.

5. The method of any of the preceding claims, wherein the insoluble solid support is selected from the group consisting of glasses, electronic devices, silicium supports, plastic supports, compact discs, filters, metallic supports or a mixture thereof.

6. The method of any of the preceding claims, wherein the length of the target nucleotide sequences to be detected and/or be quantified is determined by the use of specific primers for the retro-transcription or amplification of the 3' end of the original RNA or DNA nucleotide sequences to be detected and/or quantified in the biological sample.

7. The method of any of the preceding claims 1 to 5, wherein the length of the target nucleotide sequences to be detected and/or quantified is determined by the use of specific primer and blocking nucleotide sequence for the retro-transcription of the original RNA nucleotide sequences to be detected and/or quantified in the biological sample.

8. The method of any of the preceding claims, wherein the original nucleotide sequences to be detected and/or quantified in the biological sample are rRNAs, preferably selected from the group consisting of 16S, 23S, 18S and 28S rRNAs.

9. The method of any of the preceding claims, wherein the original nucleotide sequences to be detected and/or quantified in the biological sample are FemA specific genetic sequences of Staphylococci species, preferably the FemA genes of the strain Staphylococcus aureus, epidermidis, saprophyticus, hominis and/or haemolyticus.

10. A diagnostic and/or quantification kit or apparatus, which comprises an insoluble solid support upon which single stranded capture nucleotide sequences (allowing a specific hybridisation with target nucleotide sequences to be detected and/or quantified), are bounded, preferably by a covalent link, said single stranded capture nucleotide sequences being disposed upon the surface of the solid support according to an array with a density of at least 5 different bound single stranded capture nucleotide sequences / cm^2 surface of the solid support and in that said single stranded capture nucleotide sequences have a length comprised between about 40 and about 400 bases.

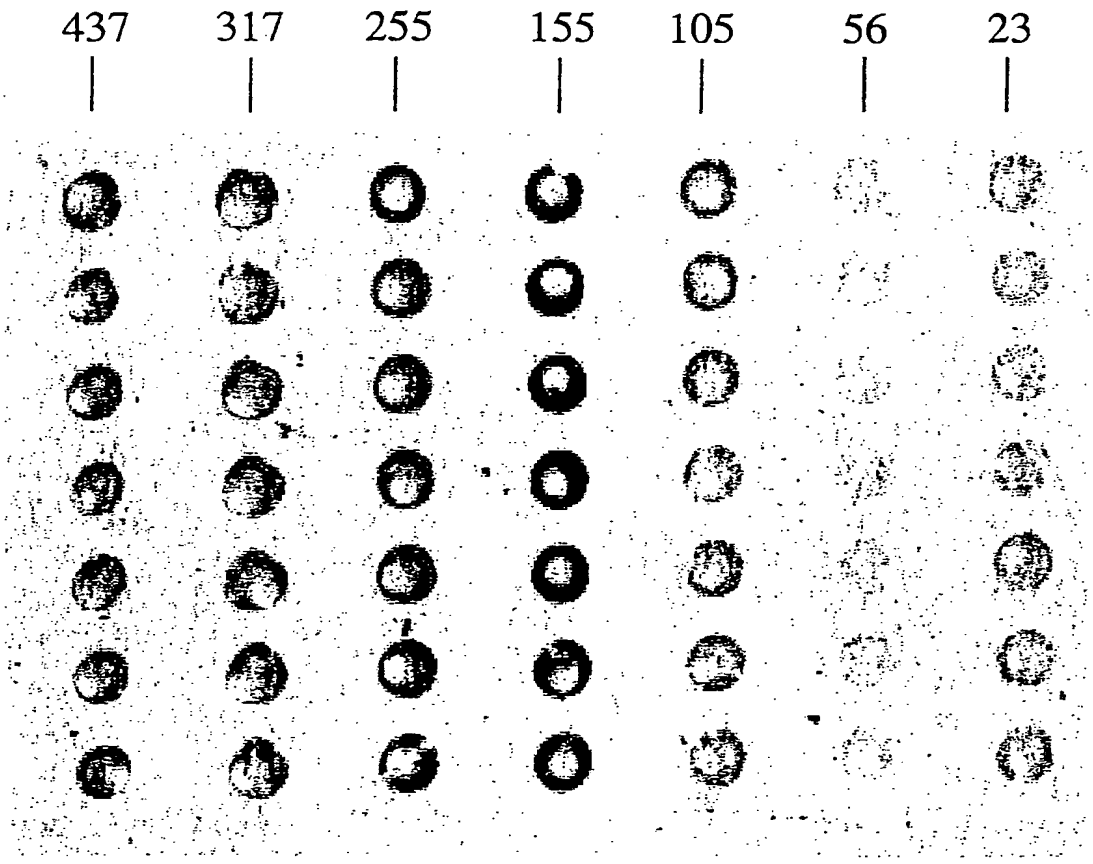
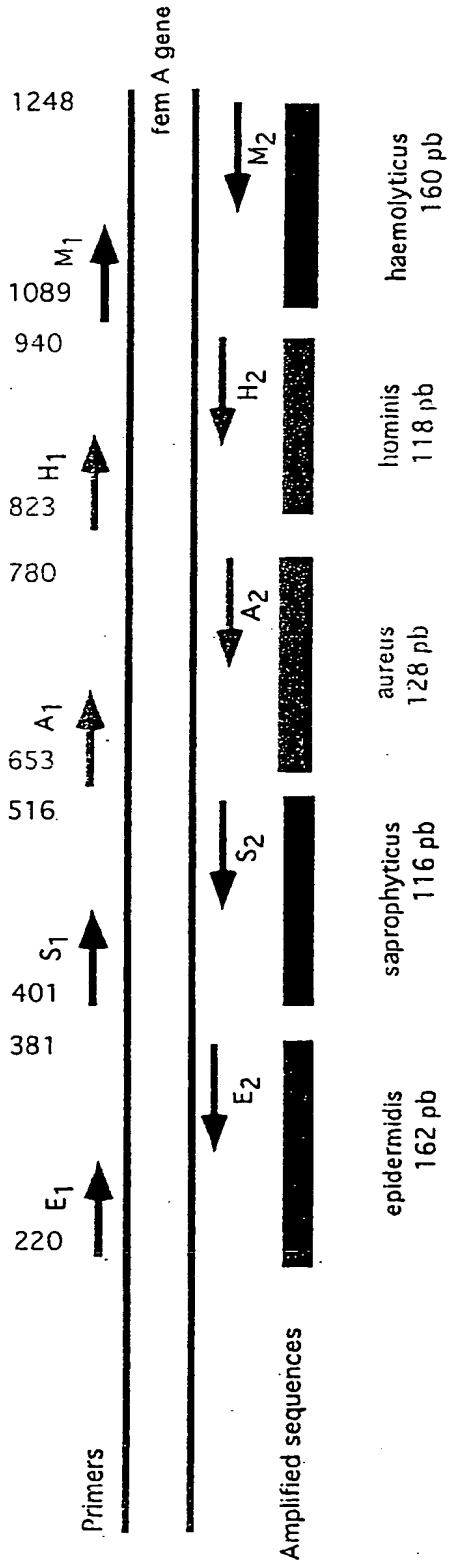
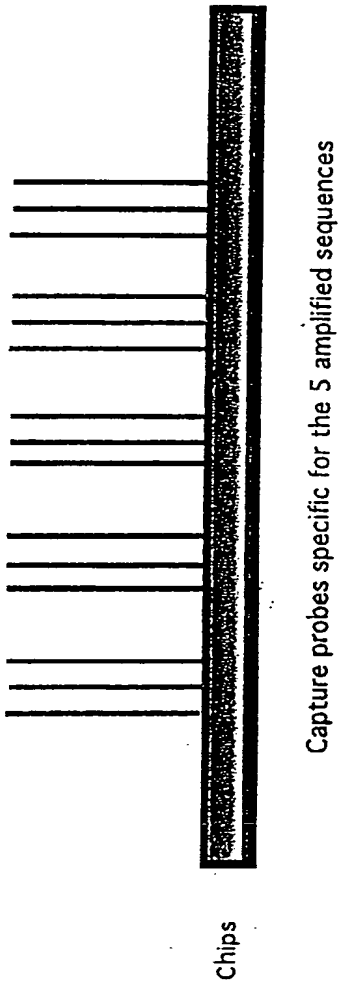
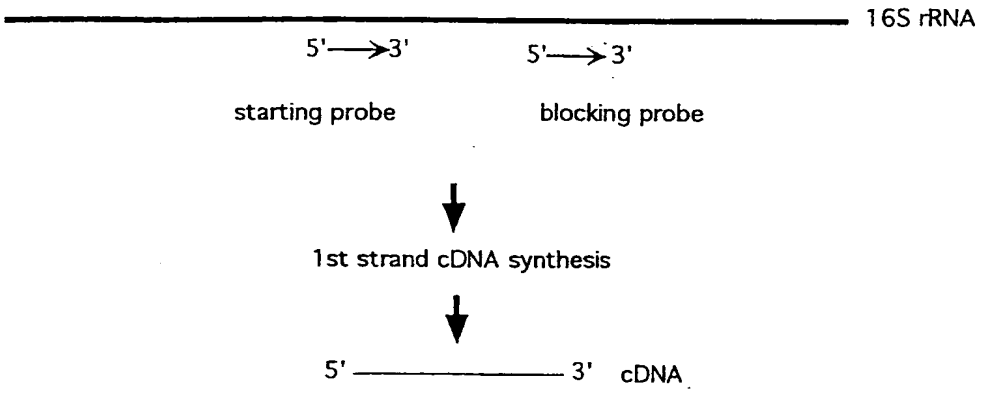


Fig. 1


fem A gene
1248
M1
1089
940
H1
823
780
A1
653
516
S1
401
381
E1
220
M2
H2
A2
S2
E2
Primers
Amplified sequences
haemolyticus
160 pb
hominis
118 pb
aureus
128 pb
saprophyticus
116 pb
epidermidis
162 pb


Chips
Capture probes specific for the 5 amplified sequences

A) Copy a specific portion of 16S rRNA



B) Copy of the 16S rRNA extremity

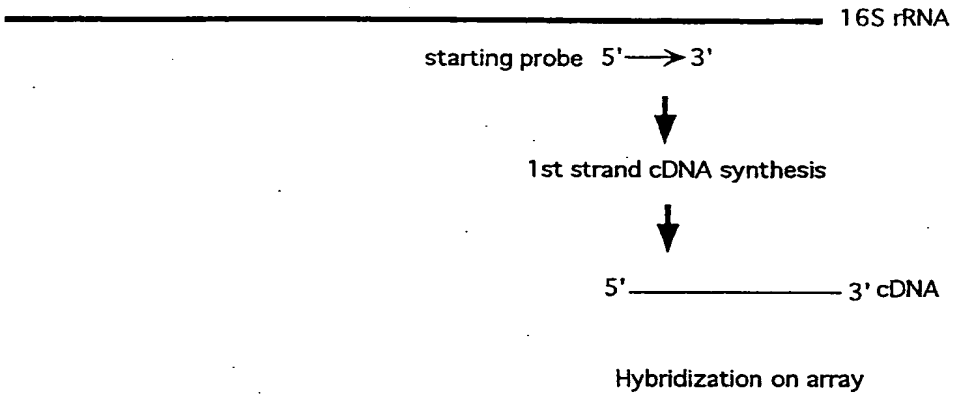


Fig. 3